

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-122811

(43)Date of publication of application : 28.04.2000

(51)Int.Cl.

G06F 3/06

G06F 12/16

G06F 13/00

(21)Application number : 10-290275

(71)Applicant : HITACHI LTD

(22)Date of filing : 13.10.1998

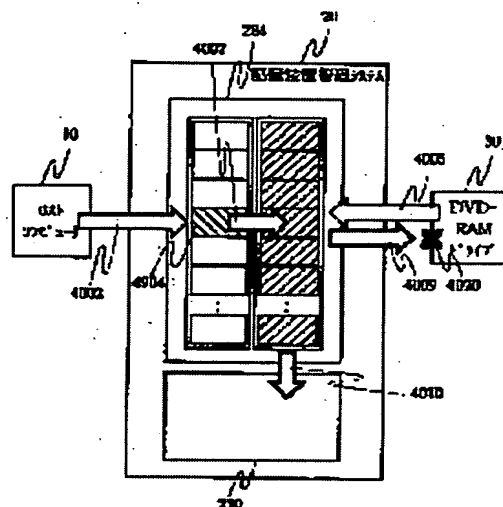
(72)Inventor : KANEDA TAISUKE

(54) STORAGE DEVICE MANAGING METHOD AND STORAGE DEVICE MANAGEMENT SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To prevent data arranged in an ECC generation unit from being entirely eliminated when a failure takes place at the time of receiving write data from an information processor, reading insufficient data in the ECC generation unit from a storage device, generating an ECC after arranging the data in the ECC generation unit and writing the data and the ECC in the storage device.

SOLUTION: When a failure takes place (4020) when write data (4904) that are smaller than an ECC generation unit are received (4002) from a host computer (10), insufficient data in the ECC generation unit are read (4005) from a DVD-RAM drive (30), an ECC is generated after the data are arranged in the ECC generation unit and the data and the ECC are written in the DVD-RAM drive (30), the data arranged in the ECC generation unit are saved in a flash memory (25). Thus, not only the write data that are smaller than the ECC generation unit but also the entire ECC block data are prevented from being eliminated.



LEGAL STATUS

[Date of request for examination]

20.02.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3616985

[Date of registration]

19.11.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-122811

(P2000-122811A)

(43) 公開日 平成12年4月28日 (2000.4.28)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード (参考)
G 0 6 F 3/06	3 0 5	G 0 6 F 3/06	3 0 5 K 5 B 0 1 8
	3 0 6		3 0 6 Z 5 B 0 6 5
12/16	3 2 0	12/16	3 2 0 F 5 B 0 8 3
13/00	3 0 1	13/00	3 0 1 R

審査請求 未請求 請求項の数 8 O L (全 14 頁)

(21) 出願番号 特願平10-290275

(22) 出願日 平成10年10月13日 (1998. 10. 13)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 兼田 泰典

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74) 代理人 100095511

弁理士 有近 紳志郎

最終頁に続く

(54) 【発明の名称】 記憶装置管理方法および記憶装置管理システム

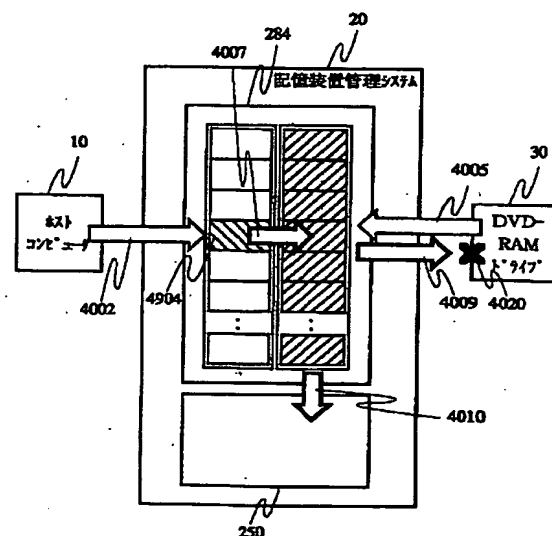
(57) 【要約】

【課題】 情報処理装置から書き込みデータを受け取り、ECC生成単位に不足するデータを記憶装置から読み出し、ECC生成単位にデータを揃えてからECCを生成し、記憶装置にデータとECCとを書き込む時に障害が発生した場合に、ECC生成単位に揃えたデータを全て消失してしまうことを防止する。

【解決手段】 ホストコンピュータ(10)からECC生成単位より小さい書き込みデータ(4904)を受け取り(4002)、ECC生成単位に不足するデータをDVD-RAMドライブ(30)から読み出し(4005)、ECC生成単位にデータを揃えてからECCを生成し、DVD-RAMドライブ(30)にデータとECCとを書き込む時に障害が発生すると(4020)、ECC生成単位に揃えたデータをフラッシュメモリ(250)に退避する。

【効果】 ECC生成単位より小さい書き込みデータの消失だけでなく、ECCブロック全体のデータの消失を防止できる。

図4



【特許請求の範囲】

【請求項1】 冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取り、冗長データ生成単位に不足するデータを記憶装置から読み出し、冗長データ生成単位にデータを揃えてから冗長データを生成し前記記憶装置に前記データと冗長データとを書き込むと共に、その書き込みが失敗した場合には、前記冗長データ生成単位に揃えたデータを不揮発性の記憶手段に保持し、情報処理装置から前記冗長データ生成単位に揃えたデータに属するデータの読み出し要求があったときは前記不揮発性の記憶手段に保持したデータを優先的に使用し、前記記憶装置からは読み出さないことを特徴とする記憶装置管理方法。

【請求項2】 前記冗長データが、ECC (Error Check Code) またはRAID (Redundant Arrays of Inexpensive Disks) におけるパリティの少なくとも一方であることを特徴とする請求項1に記載の記憶装置管理方法。

【請求項3】 冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取る書き込みデータ受取手段と、冗長データ生成単位に不足するデータを前記記憶装置から読み出すデータ読出手段と、冗長データ生成単位にデータを揃えてから冗長データを生成し前記記憶装置に前記データと冗長データとを書き込むデータ書込手段と、その書き込みが失敗した場合に前記冗長データ生成単位に揃えたデータを保持する不揮発性の記憶手段と、情報処理装置から前記冗長データ生成単位のデータに属するデータの読み出し要求があったときは前記不揮発性の記憶手段に保持したデータを優先的に使用するデータ優先使用手段とを具備したことを特徴とする記憶装置管理システム。

【請求項4】 前記不揮発性の記憶手段が、記憶装置管理システムに対して着脱可能な不揮発性メモリまたは電池付メモリであることを特徴とする請求項3に記載の記憶装置管理システム。

【請求項5】 記憶装置管理システムへの通常の電源供給が停止した時に前記書き込みデータ管理単位に揃えたデータを前記不揮発性の記憶手段に退避させるのに必要な時間だけは少なくとも記憶装置管理システムへの電源をバックアップする非常用電源装置を備えたことを特徴とする請求項3または請求項4に記載の記憶装置管理システム。

【請求項6】 複数の記憶装置との組合せでRAID (Redundant Arrays of Inexpensive Disks) を構成することを特徴とする請求項3から請求項5のいずれかに記載の記憶装置管理システム。

【請求項7】 情報処理装置と記憶装置の間に介設されることを特徴とする請求項3から請求項6のいずれかに記載の記憶装置管理システム。

【請求項8】 情報処理装置または記憶装置のいずれか

の内部に設置されることを特徴とする請求項3から請求項6のいずれかに記載の記憶装置管理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、記憶装置管理方法および記憶装置管理システムに関し、さらに詳しくは、冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取り、冗長データ生成単位に不足するデータを記憶装置から読み出し、冗長データ生成単位にデータを揃えてから冗長データを生成し、前記記憶装置に前記データと冗長データとを書き込む時に、障害が発生しても、前記冗長データ生成単位に揃えたデータを全て消失しないよう保証できる記憶装置管理方法および記憶装置管理システムに関する。

【0002】

【従来の技術】情報処理装置（ホストコンピュータ）に接続する記憶装置として、最近注目を浴びているのが、DVD-RAM (Digital Versatile Disc - Random Access Memory) 装置である。DVD-RAM装置の詳細については、東芝レビュー2, 1998, VOL. 53 NO. 2, 「特集I: 広がるDVDの世界」や、National Technical Report, 1997 JUNE, 「特集DVD」等に紹介されている。

【0003】図19に、DVD-RAMドライブ30でのデータ管理方式を示す。DVD-RAMドライブ30では、データの管理は、2KBのブロックを単位としている。一方、ECC (Error Check Code) の管理は、データの再生能力を高めるため、16ブロックを単位としている（これを、ECCブロックと呼ぶ）。

【0004】図20に示すように、ホストコンピュータは、DVD-RAM装置31に対して、2KBのブロック単位でデータを読み書きする。DVD-RAM装置31では、DVD-RAMドライブ30からの読み出しはブロック単位で行うが、DVD-RAMドライブ30への書き込みはECCブロック単位で行う。例えば、図20に示すように、ホストコンピュータがブロック4のデータAの書き込みを要求すると、DVD-RAM装置31はデータAを受け取り、バッファ29004に保持する（20001）。次に、DVD-RAM装置31は、ブロック4を含むECCブロック0（ブロック0からブロック15まで）のデータをDVD-RAMドライブ30からバッファ29500に読み出し（20005）、バッファ29500内でブロック4にデータAを結合（マージ）し、次いでECCを生成した後（20010）、バッファ29500内のECCブロック0のデータおよび生成したECCをDVD-RAMドライブ30に渡し、DVD-RAM媒体に書き込む（20020）。

【0005】

【発明が解決しようとする課題】図20に示すECCブロック0の書き込み中に障害（停電、瞬電、レーザーの

発光停止、DVD-RAM媒体の障害、DVD-RAMドライブ30の故障など）が発生すると、ECCブロック0のデータすなわちブロック0からブロック15までのデータとECCとが不一致になるため、ブロック0からブロック15までの全てのデータの書き込みに失敗したことになる。すなわち、従来のDVD-RAM装置31では、1つのブロックのデータの書き込みに失敗すると、当該ブロックを含むECCブロックのデータが全て失われてしまう問題点があった。そして、ホストコンピュータ10は、書き込みを要求したブロックのデータの書き込み失敗は失敗報告により認識できるが、そのデータ以外のどのデータが失われてしまったかは全く認識できないので、より大きな障害を引き起こしかねない問題点があった。

【0006】そこで、本発明の目的は、冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取り、冗長データ生成単位に不足するデータを記憶装置から読み出し、冗長データ生成単位にデータを揃えてから冗長データを生成し、前記記憶装置に前記データと冗長データとを書き込む時に、障害が発生しても、前記冗長データ生成単位に揃えたデータを全て消失しないよう保証できる記憶装置管理方法および記憶装置管理システムを提供することにある。

【0007】

【課題を解決するための手段】第1の観点では、本発明は、冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取り、冗長データ生成単位に不足するデータを記憶装置から読み出し、冗長データ生成単位にデータを揃えてから冗長データを生成し前記記憶装置に前記データと冗長データとを書き込むと共に、その書き込みが失敗した場合には、前記冗長データ生成単位に揃えたデータを不揮発性の記憶手段に保持し、情報処理装置から前記冗長データ生成単位に揃えたデータに属するデータの読み出し要求があったときは前記不揮発性の記憶手段に保持したデータを優先的に使用し、前記記憶装置からは読み出さないことを特徴とする記憶装置管理方法を提供する。上記構成において、前記冗長データは、例えばECCや、RAID（Redundant Arrays of Inexpensive Disks）におけるパリティである。上記第1の観点の記憶装置管理方法では、書き込みが失敗した場合には、冗長データ生成単位に揃えたデータを不揮発性の記憶手段に保持し、情報処理装置から読み出し要求があったときは不揮発性の記憶手段に保持したデータを優先的に使用する。このため、冗長データ生成単位に揃えたデータを全て消失しないよう保証できる。

【0008】第2の観点では、本発明は、冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取る書き込みデータ受取手段と、冗長データ生成単位に不足するデータを前記記憶装置から読み出すデータ読

10

20

30

40

50

長データを生成し前記記憶装置に前記データと冗長データとを書き込むデータ書込手段と、その書き込みが失敗した場合に前記冗長データ生成単位に揃えたデータを保持する不揮発性の記憶手段と、情報処理装置から前記冗長データ生成単位のデータに属するデータの読み出し要求があったときは前記不揮発性の記憶手段に保持したデータを優先的に使用するデータ優先使用手段とを具備したことを特徴とする記憶装置管理システムを提供する。上記第2の観点の記憶装置管理システムでは、前記第1の観点の記憶装置管理方法を好適に実施できる。よって、冗長データ生成単位に揃えたデータを全て消失しないよう保証できる。

【0009】第3の観点では、本発明は、前記不揮発性の記憶手段が、記憶装置管理システムに対して着脱可能な不揮発性メモリまたは電池付メモリであることを特徴とする請求項3に記載の記憶装置管理システムを提供する。上記第3の観点の記憶装置管理システムでは、記憶装置管理システムから不揮発性メモリまたは電池付メモリを外し、他の記憶装置管理システムに移すことにより、他の記憶装置管理システムでデータを回復することができる。

【0010】第4の観点では、本発明は、上記構成の記憶装置管理システムにおいて、記憶装置管理システムへの通常の電源供給が停止した時に前記書き込みデータ管理単位に揃えたデータを前記不揮発性の記憶手段に退避させるのに必要な時間だけは少なくとも記憶装置管理システムへの電源をバックアップする非常用電源装置を備えたことを特徴とする記憶装置管理システムを提供する。上記第4の観点の記憶装置管理システムでは、記憶装置管理システムへの通常の電源供給が停止した時に、冗長データ生成単位に揃えたデータを不揮発性の記憶手段に退避せうる。よって、記憶装置管理システムへの通常の電源供給が停止した場合に対しても、冗長データ生成単位に揃えたデータを全て消失しないよう保証できる。

【0011】本発明の記憶装置管理システムは、複数の記憶装置と組み合わせて、RAIDを構成してもよい。また、本発明の記憶装置管理システムは、情報処理装置と記憶装置の間に介設してもよいし、情報処理装置または記憶装置のいずれかの内部に設置してもよい。

【0012】

【発明の実施の形態】以下、図を用いて本発明の実施形態を説明する。なお、これにより本発明が限定されるものではない。

【0013】《第1の実施形態》図1に、本発明の第1の実施形態にかかる記憶装置管理システムを示す。この記憶装置管理システム20は、ホストコンピュータ10とDVD-RAMドライブ30の間に介設される。記憶装置管理システム20とホストコンピュータ10の接続はファイバチャネル15で行われ、記憶装置管理システ

ム20とDVD-RAMドライブ30の接続はSCSI (Small Computer System Interface) 25で行われている。

【0014】図2に、記憶装置管理システム20のシステム構成を示す。記憶装置管理システム20は、拡張ボードとしてファイバチャネルインタフェース210とSCSIインタフェース220とを備え、これらはPCI (Peripheral Component Interconnect) バス290に接続されている。前記PCIバス290には、プログラムおよびデータの流れを制御するチップセット286を介して、CPU280と主記憶284とが接続され、そのCPU280には2次キャッシュ282が接続されている。さらに、PCIバス290には、PCMCIA (Personal Computer Memory Card International Association) コントローラ259およびPCMCIAコネクタ255を介して、フラッシュメモリ250が接続されている。

【0015】前記主記憶284は、制御プログラムを記憶すると共に、ファイバチャネルインタフェース210を介してホストコンピュータ10から受け取った要求やデータを一時的に保持すると共にSCSIインタフェース220を介してDVD-RAMドライブ30から受け取ったデータを一時的に保持するバッファとして利用する。

【0016】図3を用いて、正常時のデータの流れを説明する。ホストコンピュータ10が一つのブロックに対してデータの書き込みを要求すると、そのブロックのデータを記憶装置管理システム20が受け取り、主記憶284内のセグメント3904に保持する(3002)。CPU280は、DVD-RAMドライブ30から前記ブロックを含むECCブロックを読み出し、主記憶284に保持する(3005)。次に、読み出したECCブロックのデータにセグメント3904のデータを結合し(3007)、ECCを生成する。そして、ECCブロックのデータとECCをDVD-RAMドライブ30に書き込む(3009)。

【0017】図4を用いて、DVD-RAMドライブ30へのデータ書き込み時に障害が発生した場合のデータの流れを説明する。ホストコンピュータ10が一つのブロックに対してデータの書き込みを要求すると、そのブロックのデータを記憶装置管理システム20が受け取り、主記憶284内のセグメント4904に保持する(4002)。CPU280は、DVD-RAMドライブ30から前記ブロックを含むECCブロックを主記憶284に読み出し、主記憶284に保持する(4005)。次に、読み出したECCブロックのデータにセグメント4904のデータを結合し(4007)、ECCを生成する。そして、ECCブロックのデータとECCをDVD-RAMドライブ30に書き込もうとする(4009)。この時に、DVD-RAMドライブ30で障

害が発生し(4020)、書き込むことができなかった場合には、主記憶284内のECCブロックのデータを、フラッシュメモリ250に書き込む(4010)。以後、ホストコンピュータ10からのデータの書き込み要求が発行されても、記憶装置管理システム20は、データの受け取りを拒否する。また、ホストコンピュータ10から前記ECCブロックに属するブロックのデータの読み出しが要求されたときは、対応するデータが主記憶284内に存在する場合はそのデータを転送し、主記憶284内に対応するデータが存在しない場合は、フラッシュメモリ250に保持したデータを転送し、DVD-RAMドライブ30からのデータの読み出しは行わない。

【0018】図5を用いて、DVD-RAMドライブ30からのデータ読み出し時に障害が発生した場合のデータの流れを説明する。ホストコンピュータ10が一つのブロックに対してデータの書き込みを要求すると、そのブロックのデータを記憶装置管理システム20が受け取り、主記憶284内のセグメント5904に保持する(5002)。CPU280は、DVD-RAMドライブ30から前記ブロックを含むECCブロックを主記憶284に読み出そうとする(5005)。この時に、DVD-RAMドライブ30で障害が発生し(5006)、前記ECCブロックのデータを読み出すことができないと、主記憶284内のセグメント5904のデータをフラッシュメモリ250に書き込む(5010)。以後、ホストコンピュータ10からのデータの書き込み要求が発行されても、記憶装置管理システム20は、データの受け取りを拒否する。また、ホストコンピュータ10から前記ブロックのデータの読み出しを要求されたときは、対応するデータが主記憶284内に存在する場合はそのデータを転送し、主記憶284内に対応するデータが存在しない場合は、フラッシュメモリ250に保持したデータを転送し、DVD-RAMドライブ30からのデータの読み出しは行わない。

【0019】DVD-RAMドライブ30で発生した障害が、DVD-RAM媒体以外の障害であり、DVD-RAM媒体には障害がない場合、障害の発生したDVD-RAMドライブ30内のDVD-RAM媒体を取り出し、別の記憶装置管理システム(20)に接続されたDVD-RAMドライブ(30)に挿入する。さらに、障害の発生したDVD-RAMドライブ30を接続していた記憶装置管理システム20からフラッシュメモリ250を外し、前記DVD-RAM媒体を挿入した記憶装置管理システム(20)にフラッシュメモリ250を接続する。この状態で記憶装置管理システム(20)を起動すると、CPU(280)はフラッシュメモリ250内にデータが保持されているか判定し、データが保持されている場合には、そのデータを前記DVD-RAM媒体に書き込む。すなわち、そのデータのブロックを含むE

CCブロックのデータをDVD-RAMドライブ(30)から読み出し、フラッシュメモリ250内のデータを結合し、ECCを生成し、前記ECCブロックのデータとECCをDVD-RAMドライブ(30)に書き込む。このように、フラッシュメモリ250とDVD-RAM媒体とを他の記憶装置管理システム(20)とDVD-RAMドライブ(30)に移せば、データを元のDVD-RAM媒体上に回復できる。

【0020】上記第1の実施形態では、主記憶284内にECCグループ2面分(ホストコンピュータ10からのデータ受け取り用とDVD-RAMドライブ30からのデータ読み出し及び結合用)のバッファを用いたが、主記憶284内に複数の領域を設けて、LRU(Least Recently Used)等のキャッシュアルゴリズムを併用してデータを管理することで、DVD-RAMドライブ30に対する読み出しおよび書き込み回数を削減し、ホストコンピュータ10からの処理要求に対してより高速に対応できるようにしてもよい。なお、複数の領域を設けた場合、DVD-RAMドライブ30に未書き込みの多くのデータが主記憶284に存在するため、これを退避できるだけの十分な容量のフラッシュメモリ250を用いることが望ましい。

【0021】フラッシュメモリ250を使用した場合、記憶装置管理システム20を含むシステム全体が停電などの障害に遭ったとき、フラッシュメモリ250へのデータの退避を実施できず、データが失われることがある。そこで、図21に示すように、電池付SRAM(Static Random Access Memory)251を、図2のフラッシュメモリ250に代えて使用してもよい。この場合、主記憶284内ではなく、電池付SRAM251内に前記バッファを設ける。これによって、システム全体が停電などの障害に遭ったときでも、データを保持しつづけることができる。障害等からの復旧後、記憶装置管理システム20が起動すると、CPU280は、電池付SRAM251内にデータが保持されているか判定し、保持されていた場合には、そのデータを前記DVD-RAM媒体に書き込む。

【0022】また、図22に示すように、無停止電源装置299から記憶装置管理システム20に電源を供給してもよい。これによって、システム全体が停電などの障害に遭ったときでも、主記憶284内のデータをフラッシュメモリ250に退避することができる。

【0023】以上の第1の実施形態によれば、DVD-RAMドライブ30へのデータの書き込みが失敗した場合でも、ECCブロックの全データが、フラッシュメモリ250または電池付SRAM251に保持されているので、DVD-RAMドライブ30での障害に伴うECCブロックの全データの消失を防ぐことができる。

【0024】《第2の実施形態》図6に、本発明の第2の実施形態にかかる記憶装置管理システムを示す。この

記憶装置管理システム21は、ホストコンピュータ10に接続されたチェンジャ装置40内に組み込まれている。チェンジャ装置40は、記憶装置管理システム21と、複数のDVD-RAMドライブ30と、DVD-RAM媒体を保持する複数の格納庫420と、それら格納庫420とDVD-RAMドライブ30の間でDVD-RAM媒体を搬送するメディア搬送ユニット450と、そのメディア搬送ユニット450を制御するロボット制御回路410とを具備している。

【0025】ホストコンピュータ10と記憶装置管理システム21はファイバチャネル15で接続され、記憶装置管理システム21とDVD-RAMドライブ30はSCSI25で接続され、記憶装置管理システム21とロボット制御回路410はRS232Cインタフェース411で接続されている。

【0026】記憶装置管理システム21がロボット制御回路410にDVD-RAM媒体の搬送を指示すると、ロボット制御回路410は、格納庫420からDVD-RAM媒体をDVD-RAMドライブ30に搬送したり、DVD-RAMドライブ30内のDVD-RAM媒体を格納庫420に戻すように、メディア搬送ユニット450を制御する。

【0027】図7に、記憶装置管理システム21のシステム構成を示す。この記憶装置管理システム21は、第1の実施形態で示した記憶装置管理システム20と基本的に同じ構成であり、データの書き込みや障害時のデータの保持の動作は第1の実施形態で説明したとおりである。それに加えて、CPU280は、ホストコンピュータ10からのDVD-RAM媒体の搬送要求を受け取り、ロボット制御インタフェース240を介してロボット制御回路410に指示し、DVD-RAM媒体の搬送を実行する。

【0028】上記第2の実施形態においても、第1の実施形態と同様に、キャッシュアルゴリズムを併用したり、電池付SRAMを用いたり、無停止電源装置を使用してもよい。

【0029】以上の第2の実施形態によれば、第1の実施形態の効果に加えて、記憶装置管理システム21をチェンジャ装置40の一部として実現することで、記憶装置管理システム21の実施コストを低減することができる。

【0030】《第3の実施形態》図8に、本発明の第3の実施形態にかかる記憶装置管理システムを示す。この記憶装置管理システム22は、ホストコンピュータ10と4台のチェンジャ装置80の間に接続されている。ホストコンピュータ10と記憶装置管理システム22はファイバチャネル15で接続し、記憶装置管理システム22とチェンジャ装置80はSCSI25で接続している。

【0031】図9に、チェンジャ装置80の構成を示

す。このチェンジャ装置80は、DVD-RAM媒体を保持する複数の格納庫420と、それら格納庫420とDVD-RAMドライブ30の間でDVD-RAM媒体を搬送するメディア搬送ユニット450と、そのメディア搬送ユニット450を制御するロボット制御回路410と、記憶装置管理システム22からのDVD-RAM媒体の搬送要求を受け取ってロボット制御回路410への命令に変換するチェンジャSCSIインタフェース405とを具備している。チェンジャ装置80内のDVD-RAMドライブ30は、SCSI25で、記憶装置管理システム22に接続している。

【0032】図10に、記憶装置管理システム22のシステム構成を示す。この記憶装置管理システム22は、第1の実施形態で示した記憶装置管理システム20と基本的に同じ構成であるが、それに加えて、4台のチェンジャ装置80にそれぞれ接続するためのSCSIインタフェース220を4つ具備している。これにより、4台のチェンジャ装置80内の各DVD-RAMドライブ30を組み合わせて、RAID構成を実現する。すなわち、4台のDVD-RAMドライブ30のうち3台をデータ格納ドライブとし、1台をRAIDにおけるパリティ格納ドライブとする。また、CPU280は、RAID処理を実現するため、データを3台のデータ格納ドライブに分散するデータ分散モジュールと、3台のデータ格納ドライブに分散したデータからパリティを生成するパリティ生成モジュールと、ホストコンピュータ10から読み出し要求があったときに3台のデータ格納ドライブから読み出したデータを集めてホストコンピュータ10に転送するデータ集合モジュールと、1台のデータ格納ドライブが故障したときに残りの2台のデータ格納ドライブのデータとパリティ格納ドライブのパリティから故障したデータ格納ドライブのデータを復元するデータ復元モジュールとを実装している。

【0033】図11を用いて、正常時のデータの流れを説明する。ホストコンピュータ10が一つの論理ブロックに対するデータの書き込みを記憶装置管理システム22に要求すると、CPU280のデータ分散モジュールが、前記データに対応するDVD-RAMドライブと当該DVD-RAMドライブにおける物理ブロックを算出し、ホストコンピュータ10からデータを受け取り、主記憶284内のセグメント11936に保持する(11002)。次に、CPU280は、前記物理ブロックを含むECCブロックのデータを前記データに対応するDVD-RAMドライブ30から読み出す(11023)と共に、そのECCブロックを含むRAIDグループのデータ(一つのパリティを生成するのに使用する一群のECCブロックのデータ)を他のDVD-RAMドライブ30から主記憶284に読み出す(11021、11022)。次に、セグメント11936のデータを、対応するDVD-RAMドライブ30から読み出したデータ

に結合する(11070)。そして、ECCを生成し、新たなECCブロックとする。次に、CPU280のパリティ生成モジュールが、新たなECCブロックと他のデータを用いてパリティを生成する。次に、新たなECCブロックのデータを、対応するデータ格納ドライブに書き込む(11083)と共に、パリティをパリティ格納ドライブに書き込む(11084)。

【0034】図12～図13を用いて、1台のDVD-RAMドライブ30へのデータの書き込み時に障害が発生した場合のデータの流れを説明する。図12において、ホストコンピュータ10が一つの論理ブロックに対するデータの書き込みを記憶装置管理システム22に要求すると、CPU280のデータ分散モジュールが、前記データに対応するDVD-RAMドライブと当該DVD-RAMドライブにおける物理ブロックを算出し、ホストコンピュータ10からデータを受け取り、主記憶284内のセグメント12936に保持する(12002)。次に、CPU280は、前記物理ブロックを含むECCブロックのデータを前記データに対応するDVD-RAMドライブ30から読み出す(12023)と共に、そのECCブロックを含むRAIDグループのデータを他のDVD-RAMドライブ30から主記憶284に読み出す(12021、12022)。次に、セグメント12936のデータを、対応するDVD-RAMドライブ30から読み出したデータに結合する(12070)。そして、ECCを生成し、新たなECCブロックとする。次に、CPU280のパリティ生成モジュールが、新たなECCブロックと他のデータを用いてパリティを生成する。次に、新たなECCブロックのデータを、対応するデータ格納ドライブに書き込もうとする(12083)と共に、パリティをパリティ格納ドライブに書き込もうとする(12084)。この時、データを書き込もうとしたデータ格納ドライブで障害が発生し(12006)、書き込むことができなかったとする。【0035】すると、図13に示すように、書き込むことができなかった新たなECCブロックのデータを、他のすべてのDVD-RAMドライブ30の障害データ保持領域に書き込む(13011、13012、13014)。これらの障害データ保持領域は、通常のデータもしくはパリティを書き込む領域とは別に設けた特別な領域である。以後、ホストコンピュータ10からデータの書き込みが要求されても、記憶装置管理システム22は、データの受け取りを拒否する。また、ホストコンピュータ10から書き込むことができなかった新たなECCブロックのデータの読み出しを要求された場合には、そのデータが主記憶284内に存在すればそれを転送し、主記憶284内に対応するデータが存在しないなら障害データ保持領域に保持しておいた当該データを読み出して転送し、障害を発生したデータ格納ドライブからのデータの読み出しは行わない。

【0036】図14を用いて、全てのDVD-RAMドライブ30が使用できなくなった場合のデータの流れを説明する。全てのDVD-RAMドライブ30が使用できなくなった場合、書き込むことができなかった新たなECCブロックのデータの障害データ保持領域への書き込みも失敗する(14011、14012、14014)。そこで、CPU280は、書き込むことができなかった新たなECCブロックのデータを、フラッシュメモリ250に退避する(14050)。以後、ホストコンピュータ10からデータの書き込みが要求されても、記憶装置管理システム22は、データの受け取りを拒否する。また、ホストコンピュータ10から、書き込むことができなかった新たなECCブロックのデータの読み出しを要求された場合には、そのデータが主記憶284内に存在すればそれを転送し、主記憶284内に対応するデータが存在しないならフラッシュメモリ250に退避しておいた当該データを読み出して転送し、障害が発生したデータ格納ドライブからのデータの読み出しは行わない。

【0037】上記第3の実施形態においても、第1の実施形態と同様に、キャッシュアルゴリズムを併用したり、電池付SRAMを用いたり、無停止電源装置を使用してもよい。

【0038】以上の第3の実施形態によれば、RAID構成をとる場合においても、第1の実施形態の効果を得ることができる。

【0039】《第4の実施形態》図15に、本発明の第4の実施形態にかかる記憶装置管理システムを示す。この記憶装置管理システムは、DVD-RAMドライブ30をSCSI25で接続したホストコンピュータ10内の記憶装置管理デバイスドライバ90として実現されている。この記憶装置管理デバイスドライバ90の制御下に、DVD-RAMドライブ30と、PCMCIAコネクタ255およびフラッシュメモリ250を置く。

【0040】図16に、記憶装置管理デバイスドライバ90の内部構成を示す。記憶装置管理デバイスドライバ90は、ホストコンピュータ10のファイルシステムソフトウェア(図17の11)とのインタフェースを実現するファイルシステムインタフェースソフトウェアモジュール910と、ホストコンピュータ10内の主記憶に設けた2面のバッファ951と952を管理するバッファ制御ソフトウェアモジュール920と、DVD-RAMドライブ30を制御するDVD-RAMドライブ制御ソフトウェアモジュール960と、SCSIインタフェースを制御するSCSIインタフェース制御ソフトウェアモジュール970と、フラッシュメモリ250を制御するフラッシュメモリ制御ソフトウェアモジュール980とを具備している。

【0041】図17を用いて、正常時のデータの流れを説明する。ホストコンピュータ10のファイルシステム

ソフトウェア11が一つのブロックのデータの書き込みを要求すると、ファイルシステムインタフェースソフトウェアモジュール910がそのデータを受け取り(17002)、バッファ951内のセグメント17904に保持する(17005)。バッファ制御ソフトウェアモジュール920は、DVD-RAMドライブ制御ソフトウェアモジュール960とSCSIインタフェース制御ソフトウェアモジュール970を介して、DVD-RAMドライブ30から前記ブロックを含むECCブロックをバッファ952に読み出す(17006)。次に、読み出したECCブロックのデータにセグメント17904のデータを結合する(17007)。次いで、ECCブロックのデータとECCをDVD-RAMドライブ30に書き込む(17010)。

【0042】図18を用いて、DVD-RAMドライブ30への書き込み時に障害が発生した場合を説明する。上記DVD-RAMドライブ30への書き込み(17010)時に障害が発生し、データを書き込むことができなかった場合、バッファ制御ソフトウェアモジュール920は、フラッシュメモリ制御ソフトウェアモジュール980を介して、バッファ952内のECCブロックのデータをフラッシュメモリ250に書き込む(18030)。以後、ホストコンピュータ10のファイルシステムソフトウェア11からデータの書き込み要求が発行されても、バッファ制御ソフトウェアモジュール920は、データの受け取りを拒否する。また、ホストコンピュータ10のファイルシステムソフトウェア11から前記ECCブロック内のブロックのデータの読み出しを要求されたときには、そのデータがバッファ952内に存在する場合はそのデータを転送し、バッファ952内に存在しない場合はフラッシュメモリ250に保持したデータを転送し、DVD-RAMドライブ30からのデータの読み出しは行わない。

【0043】上記第4の実施形態においても、第1の実施形態と同様に、キャッシュアルゴリズムを併用したり、電池付SRAMを用いたり、無停止電源装置を使用してもよい。

【0044】以上の第4の実施形態によれば、ホストコンピュータ10内に記憶装置管理デバイスドライバ90を設けるので、簡単な構成で、第1の実施形態の効果を得ることができる。

【0045】

【発明の効果】本発明の記憶装置管理方法および記憶装置管理システムによれば、冗長データ生成単位より小さい書き込みデータを情報処理装置から受け取り、冗長データ生成単位に不足するデータを記憶装置から読み出し、冗長データ生成単位にデータを揃えてから冗長データを生成し、前記記憶装置に前記データと冗長データとを書き込む時に、障害が発生しても、前記冗長データ生成単位に揃えたデータを全て消失しないよう保証でき

る。

【図面の簡単な説明】

【図1】第1の実施形態における記憶装置管理システムの接続を示す図である。

【図2】第1の実施形態における記憶装置管理システムの内部構成を示す図である。

【図3】第1の実施形態における正常時のデータの流れを示す図である。

【図4】第1の実施形態におけるDVD-RAMドライブでの書き込み障害による書き込み失敗時のデータの流
10

【図5】第1の実施形態におけるDVD-RAMドライブでの読み出し障害による書き込み失敗時のデータの流
れを示す図である。

【図6】第2の実施形態における記憶装置管理システムの接続を示す図である。

【図7】第2の実施形態における記憶装置管理システムの内部構成を示す図である。

【図8】第3の実施形態における記憶装置管理システムの接続を示す図である。

【図9】第3の実施形態におけるチェンジャ装置の内部構成を示す図である。

【図10】第3の実施形態における記憶装置管理システムの内部構成を示す図である。

【図11】第3の実施形態における正常時のデータの流れを示す図である。

【図12】第3の実施形態における1台のDVD-RAMドライブへの書き込み障害による書き込み失敗時のデータの流
れの前半を示す図である。

【図13】第3の実施形態における1台のDVD-RAMドライブへの書き込み障害による書き込み失敗時のデータの流
れの後半を示す図である。

【図14】第3の実施形態における全てのDVD-RAMドライブへの書き込み障害による書き込み失敗時のデータの流
れを示す図である。

【図15】第4の実施形態における記憶装置管理システム（デバイスドライバ）の接続を示す図である。

【図16】第4の実施形態における記憶装置管理システム

*ム（デバイスドライバ）の内部構成を示す図である。

【図17】第4の実施形態における正常時のデータの流れを示す図である。

【図18】第4の実施形態におけるDVD-RAMドライブへの書き込み障害による書き込み失敗時のデータの流
れを示す図である。

【図19】DVD-RAMドライブにおけるデータの管理単位（ブロック）とECCの生成単位（ECCブロック）を示す図である。

【図20】従来のDVD-RAM装置におけるデータの流れを示す図である。

【図21】第1の実施形態の記憶装置管理システムの第1の変形例の内部構成を示す図である。

【図22】第1の実施形態の記憶装置管理システムの第2の変形例の構成を示す図である。

【符号の説明】

10…ホストコンピュータ、11…ファイルシステムソフトウェア、15…ファイバチャネル、20…記憶装置管理システム、21…記憶装置管理システム、22…記憶装置管理システム、25…SCSI、30…DVD-RAMドライブ、31…DVD-RAM装置、40…チェンジャ装置、80…チェンジャ装置、90…記憶装置管理デバイスドライバ、210…ファイバチャネルインタフェース、220…SCSIインタフェース、240…ロボット制御インタフェース、250…フラッシュメモリ、251…電池付SRAM、255…PCMCIAコネクタ、259…PCMCIAコントローラ、280…CPU、282…2次キャッシュ、284…主記憶、286…チップセット、299…無停止電源装置、405…チェンジャSCSIインタフェース、410…ロボット制御回路、420…格納庫、450…メディア搬送ユニット、910…ファイルシステムインタフェースコントローラ、920…バッファ制御ソフトウェアモジュール、951…バッファ、952…バッファ、960…DVD-RAMドライブ制御ソフトウェアモジュール、970…SCSIインタフェース制御ソフトウェアモジュール、980…フラッシュメモリ制御ソフトウェアモジュール。

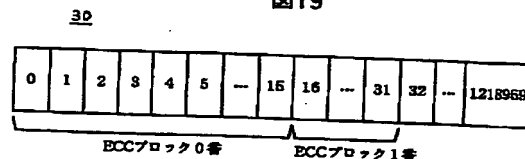
【図1】

図1



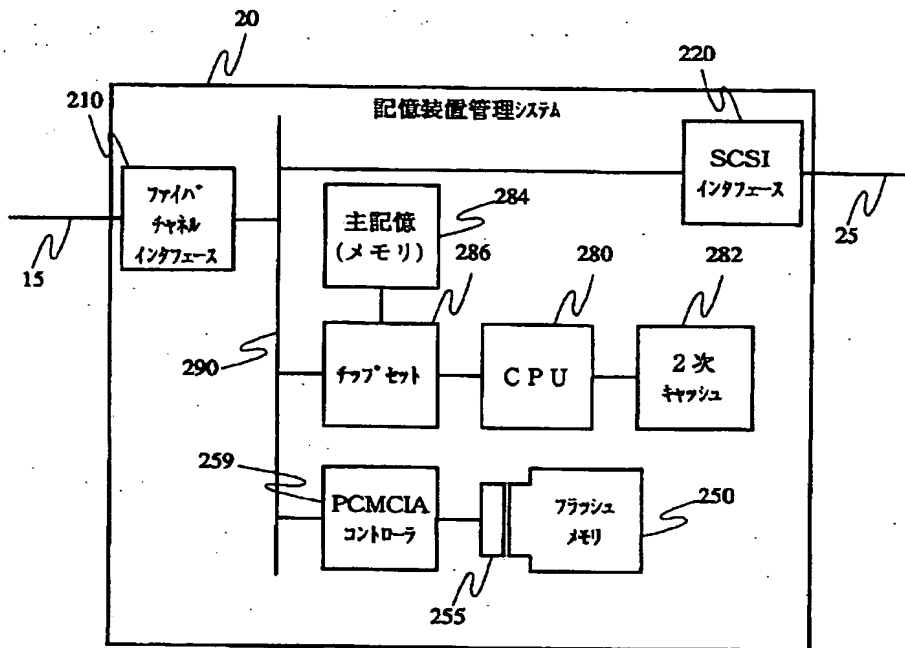
【図19】

図19



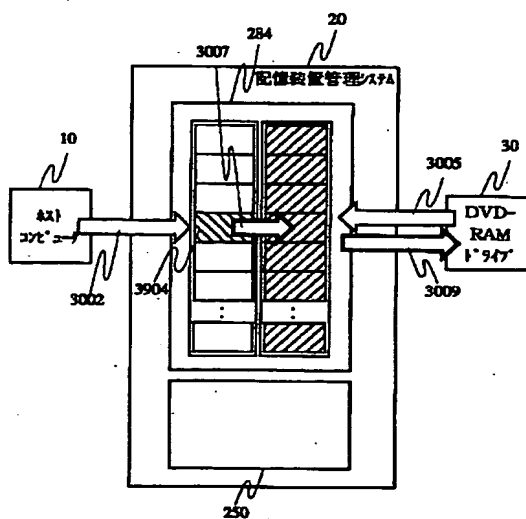
【図2】

図2



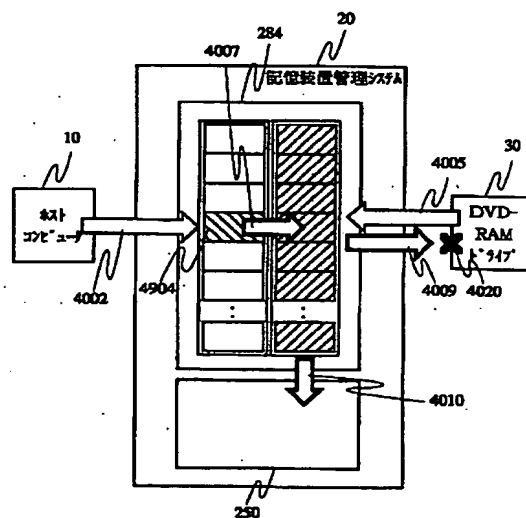
【図3】

図3

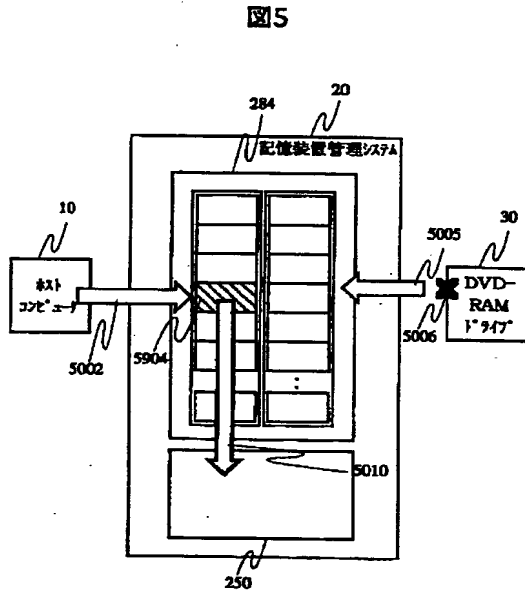


【図4】

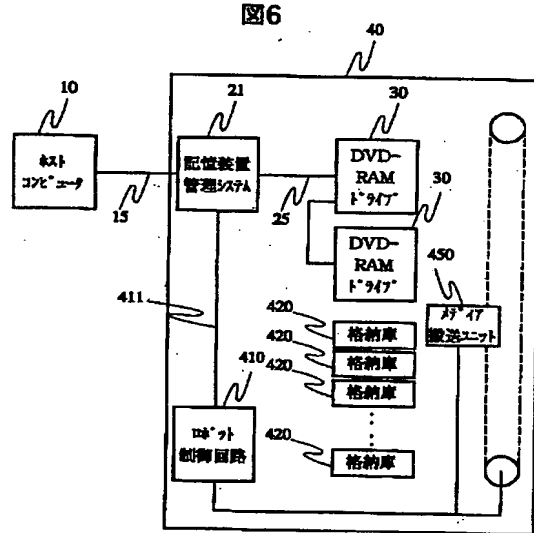
図4



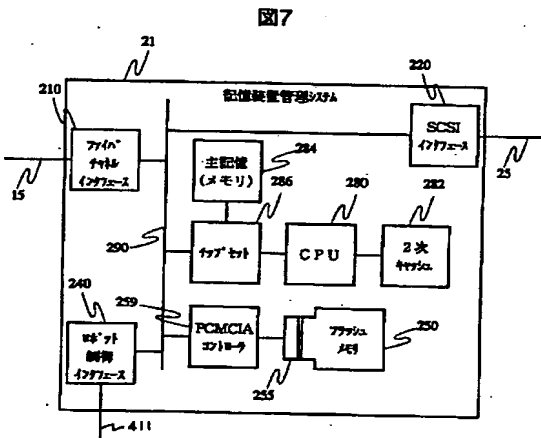
【図5】



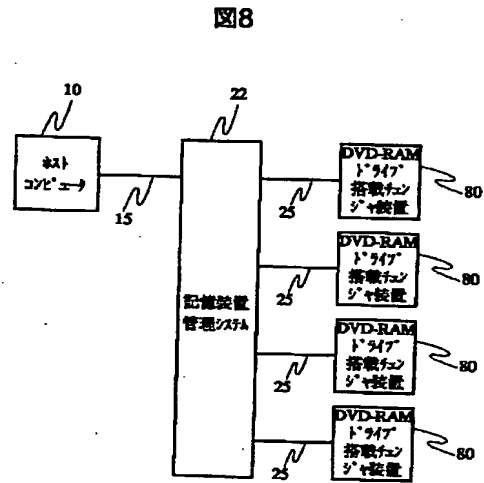
【図6】



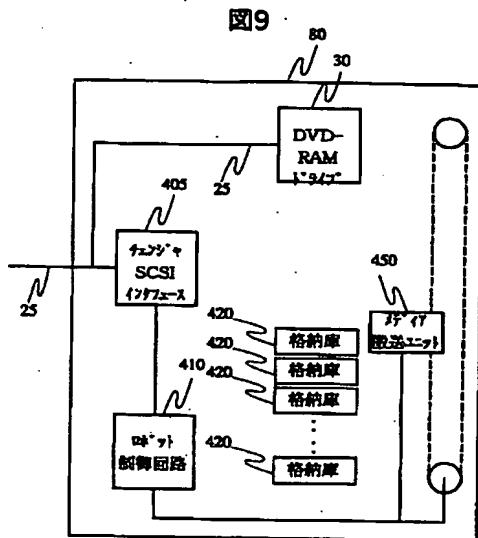
【図7】



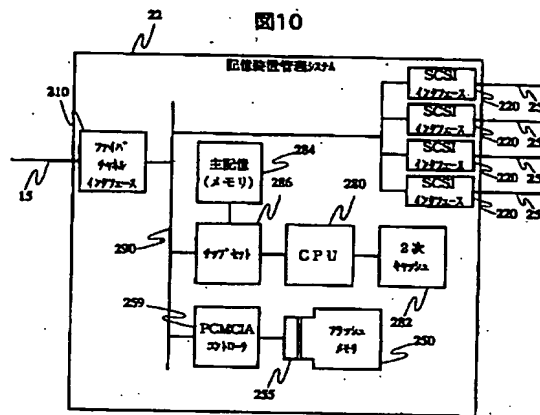
【図8】



【図9】



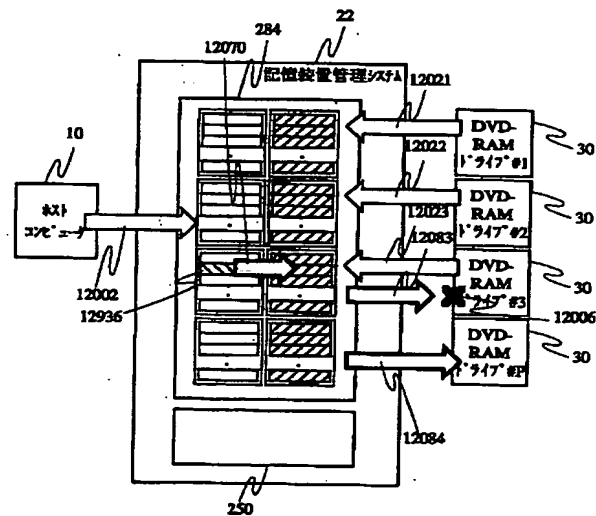
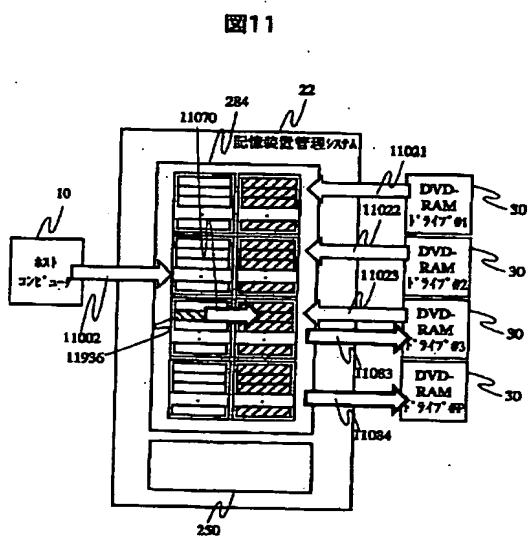
【図10】



【図12】

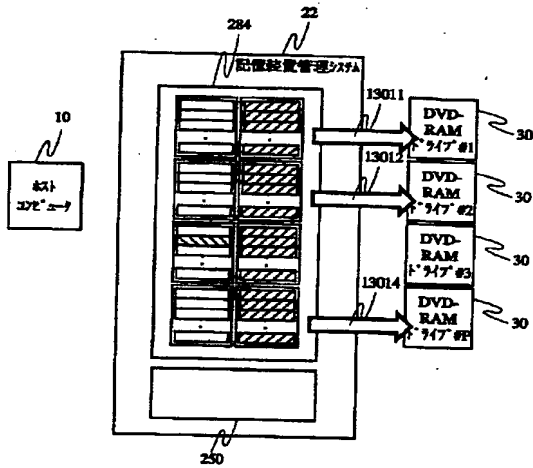
図12

【図11】



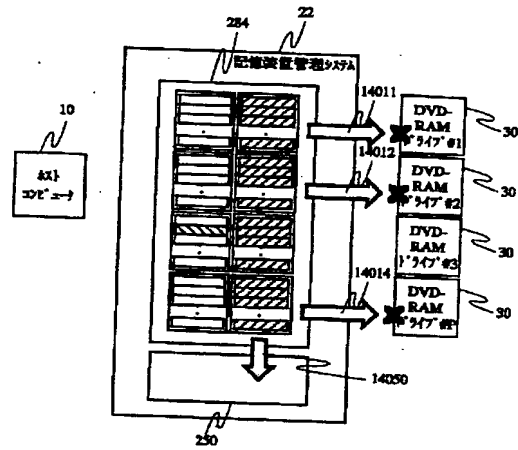
【図13】

図13



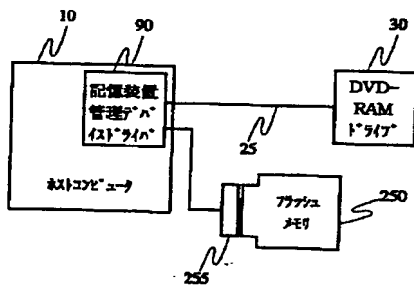
【図14】

図14



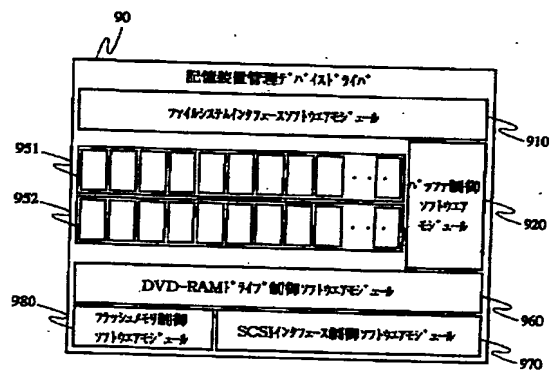
【図15】

図15

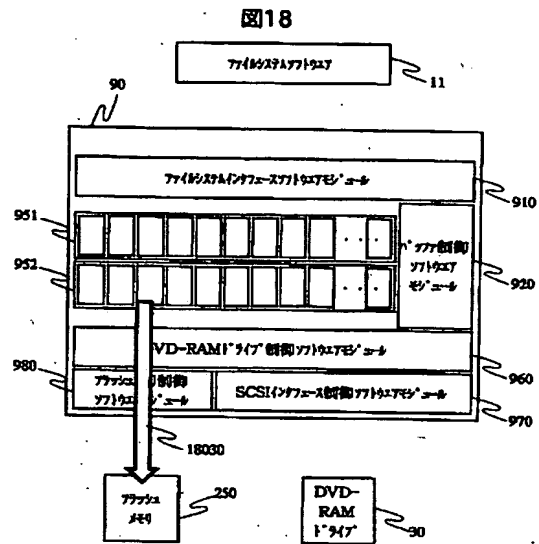


【図16】

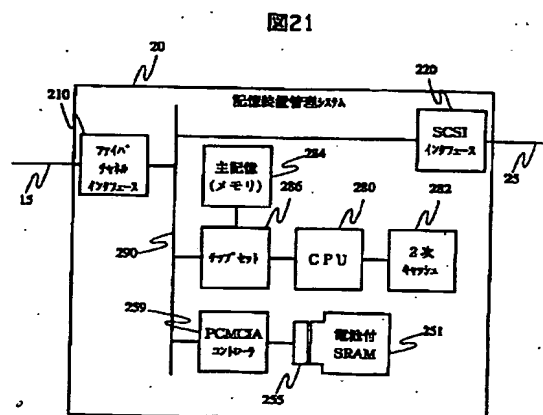
図16



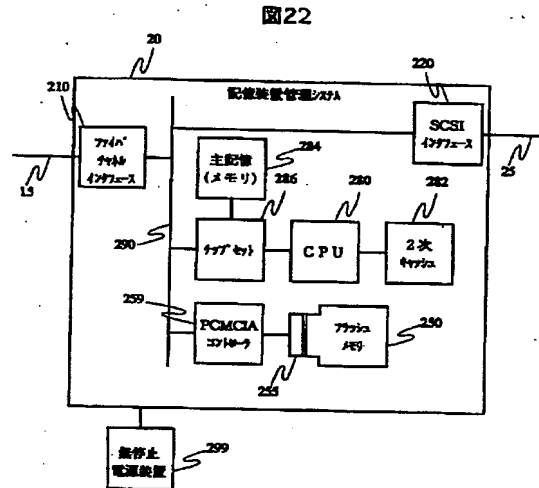
【圖 18】



【圖 2 1】



【図22】



フロントページの続き

F ターム(参考) 5B018 GA04 HA03 KA14 LA01 MA16
 NA06 QA05 QA15
 5B065 BA04 CA03 CA11 CA30 CE01
 EA03 EA12 EA23 EA24 EA26
 EA39 ZA14
 5B083 AA09 BB01 BB03 CC02 CD13
 DD04 DD07 EE08